AD-A114 304   PRINCETON UNIV NJ DEPT OF STATISTICS   F/G 12/1
PRE-CONFIGURAL POLYSAMPLING PUSHBACK PERFORMANCE.(U)
DEC 81 K B KRYSTINIK   DAAG29-79-C-0205
UNCLASSIFIED TR-210-SER-2   ARO-16669.13-M   NL

1 OF 1
AD
A114304

END
DATE
FILMED
6-82
DTIC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(12)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 16669.13-M	2. GOVT ACCESSION NO.: AD-A114
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Pre-configural Polysampling Pushback Performance	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Katherine Bell Krystinik	8. CONTRACT OR GRANT NUMBER(s): DAAG29 79 C 0205
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Princeton University, Princeton, NJ 08540	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709	12. REPORT DATE: Dec 81
	13. NUMBER OF PAGES: 43
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE APR 26 1982 D H

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

sampling (statistics)
robust analysis
statistics
estimating

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The pushback estimates are defined by a preliminary data modification followed by the application of a robust statistic to the modified data. The form of the pushback estimate depends upon the choice of a scale estimate for the original data, a set of central values of order statistics, and a constant multiplier. Particular forms of the pushback estimates are both simple and perform rather well relative to a good estimate in the w-estimate or M-estimate category.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DTIC FILE COPY

Pre-configural polysampling pushback performance*

by

Katherine Bell Krystinik

Technical Report No. 210, Series 2
Department of Statistics
Princeton University
December 1981

*Prepared in connection with research at Princeton University, supported by the Army Research Office Durham).

ABSTRACT

The pushback estimates are defined by a prelininary data modification followed by the application of a robust statistic to the modified data. The form of the pushback estimate depends upon the choice of a scale estimate for the original data, a set of central values of order statistics, and a constant multiplier. Particular forms of the pushback estimates are both simple and perform rather well relative to a good estimate in the w-estimate or M-estimate category.

1. Introduction.

We have some quite good estimates of location in the w-estimate and M-estimate category. The pushback estimates are another category of estimates -- one that appears very different -- that satisfies two requirements. First, the pushback estimates of a certain form perform well over a wide range of distributions. Second, the estimates are relatively simple to understand and use.

Monte Carlo studies indicate that a well-chosen form of the pushback achieves a maximin efficiency (relative to the w6-biweight) of 89% for sample size 20. This performance is measured over the set of distributions including the Gaussian, one wild Gaussian, mixture, slacu, and slash.

2. The pushback.

The pushback procedures are based on preliminary data modification. The order statistics of the data are modified before a simple robust estimate of location is applied. More formally our procedure is as follows; Suppose we are given n observations,

$$y_1, y_2, \ldots, y_n ,$$

from a particular situation $\{f_i: i=1, \ldots, n\}$ where the f_i are location-scale densities. The situation may be either simple or compound (Bruce, Pregibon, Tukey (1981)). The procedure modifies the order statistics of the n observations

$$y(1), y(2), \ldots, y(n) ,$$

by subtracting some function of i, p(i),

$$y(1)-p(1), y(2)-p(2), \ldots, y(n)-p(n) .$$

The form of p(i) considered is:

$$p(i) = k \cdot s \cdot a(i)$$

where k is a constant, s is an estimate of the scale of the data {y(i)} and {a(i)} is a set of central values of order statistics from a suitable unit distribution. We then apply T, a robust estimate, to the set $\{y(i)-k \cdot s \cdot a(i)\}$ to determine a location estimate for the distribution of the {y(i)}.

We will call this procedure the pushback T when the estimate T is applied to the modified data, or pushback when we have no particular estimate in mind. The pushback was previously studied by L. Nanni (Nanni (1979)).

3. Simulation Cases.

Various forms of

- the location estimate
- the central order-statistic values
- the scale estimate

- the constant multiplier

were tested using data from a wide range of distributions representing the extremes of data distributions which are likely to be encountered in practice.

In order to preserve the simplicity of the procedure, the median and two forms of the biweight were used as the location estimates. Biweights discussed here are specified by the c value and by the number of iterations. The two choices are the w6-biweight (one iteration and c=6) and the 9-biweight (iteration to tolerance and c=9). Preliminary simulations indicated these two forms as the two with the best performance for the distributions of interest. The biweights are defined iteratively as follows:

$$T_j = \frac{\sum_i w_i y_i}{\sum_i w_i}$$

$$\text{where } w_i = \begin{cases} \left[1 - \left|\dfrac{y_i - T_{j-1}}{cs_j}\right|^2\right]^2 & \text{if } \left|\dfrac{y_i - T_{j-1}}{cs_j}\right| < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$s_j = \text{med}\{|y_i - T_{j-1}|\}\ .$$

and the initial estimate, T_o, is the median.

The subscript j denotes the number of the iteration. Iteration to tolerance is stopped when $|T_j - T_{j-1}| \leq .0005$ or when $j=20$, whichever occurs first.

Central values of order statistics for a sample of size n from a Gaussian, logistic, and a cosine-bell distribution, each with mean zero and variance one, were used. For the Gaussian and logistic distributions, expected values of order statistics were chosen as the central value ((Owen (1962)) and (Birnbaum and Dudman (1963)). For the cosine-bell density,

$$\begin{cases} \frac{1}{2}\,\frac{\sqrt{\pi^2-8}}{2}\cos\left(y\frac{\sqrt{\pi^2-8}}{2}\right) & \text{on } \left[\frac{-\pi}{\sqrt{\pi^2-8}}, \frac{\pi}{\sqrt{\pi^2-8}}\right] \\ 0 & \text{otherwise} \end{cases}$$

the values $F^{-1}\left(\frac{i-\frac{1}{3}}{n+\frac{1}{3}}\right)$ were used. $F^{-1}\left(\frac{i-\frac{1}{3}}{n+\frac{1}{3}}\right)$ has been shown to be a good approximation to the median of y(i). Letting $a(i)_{Gau}$, $a(i)_{log}$, and $a(i)_{cob}$ denote the central value of the i^{th} order statistic from a Gaussian, logistic, and cosine-bell distribution, respectively, and noting that $a(i) = -a(n+i-1)$ because of symmetry, we see in table 1 that

Table 1

Central Values of Order Statistics* for a Sample of Size 20
(Note $a(i) = a(21-i)$ for $i=1,10$)

	Gaussian	logistic	cosine-bell
a(20)	1.8675	1.95597	1.76496
a(19)	1.4076	1.37562	1.44818
a(18)	1.1309	1.06933	1.21355
a(17)	.9210	.85312	1.01459
a(16)	.7454	.68083	.83609
a(15)	.5903	.53381	.67068
a(14)	.4483	.40254	.51407
a(13)	.3149	.28137	.36341
a(12)	.1870	.16651	.21660
a(11)	.0620	.05513	.07197

*Gaussian and logistic values are the expected values of the order statistics; cosine-bell values are $F^{-1}((i-1/3)/(n+1/3))$.

$$|a(i)_{Gau}| < |a(i)_{cob}| \qquad i=2,\ \ldots,\ 19$$
$$|a(i)_{Gau}| > |a(i)_{log}|$$

$$|a(i)_{Gau}| > |a(i)_{cob}| \qquad i=1,20$$
$$|a(i)_{Gau}| < |a(i)_{log}|.$$

The differences between the Gaussian $\{a(i)\}$ and the two other sets of $\{a(i)\}$ can be seen in the plot of log $a_Q(j)$ against log $a_{Gau}(i)$ (figure 1) for Q={log,cob}.

Many scale estimates were tested. These included the median jump ratio, the median absolute deviation, and various other percentage points of $\{|y(i)-\text{med}\{y_i\}|\}$. The median jump ratio (MJR) is defined as

$$\underset{i=1,\ldots,\frac{n}{2}}{\text{med}} \left\{\frac{y(n-i+1)-y(i)}{a(n-i+1)-a(i)}\right\}$$

when n is even, as is the case for the simulations discussed here. The MJR used the same $\{a(i)\}$ as are used in pushing back the data and is thus matched to the specific pushback form. The median absolute deviation from the median (MAD) is defined as $\text{med}\{|y(i)-\text{med}\{y(i)\}|\}$, i.e., the 50% point of $\{|y(i)-\text{med}\{y(i)\}|\}$. Other percent points of $\{|y(i) - \text{med}(y(i))|\}$ were used. The complete set of lower percentage points was P= 37.5,45,50,55,70,75,80,85, and 90. For example, the 45% point is z(9) where $\{z(i)\}$ are the ordered set $\{|y(i)-\text{med}\{y(i)\}|\}$. We will refer to P% point of $\{|y(i)-\text{med}\{y(i)\}|\}$ as P%AD.

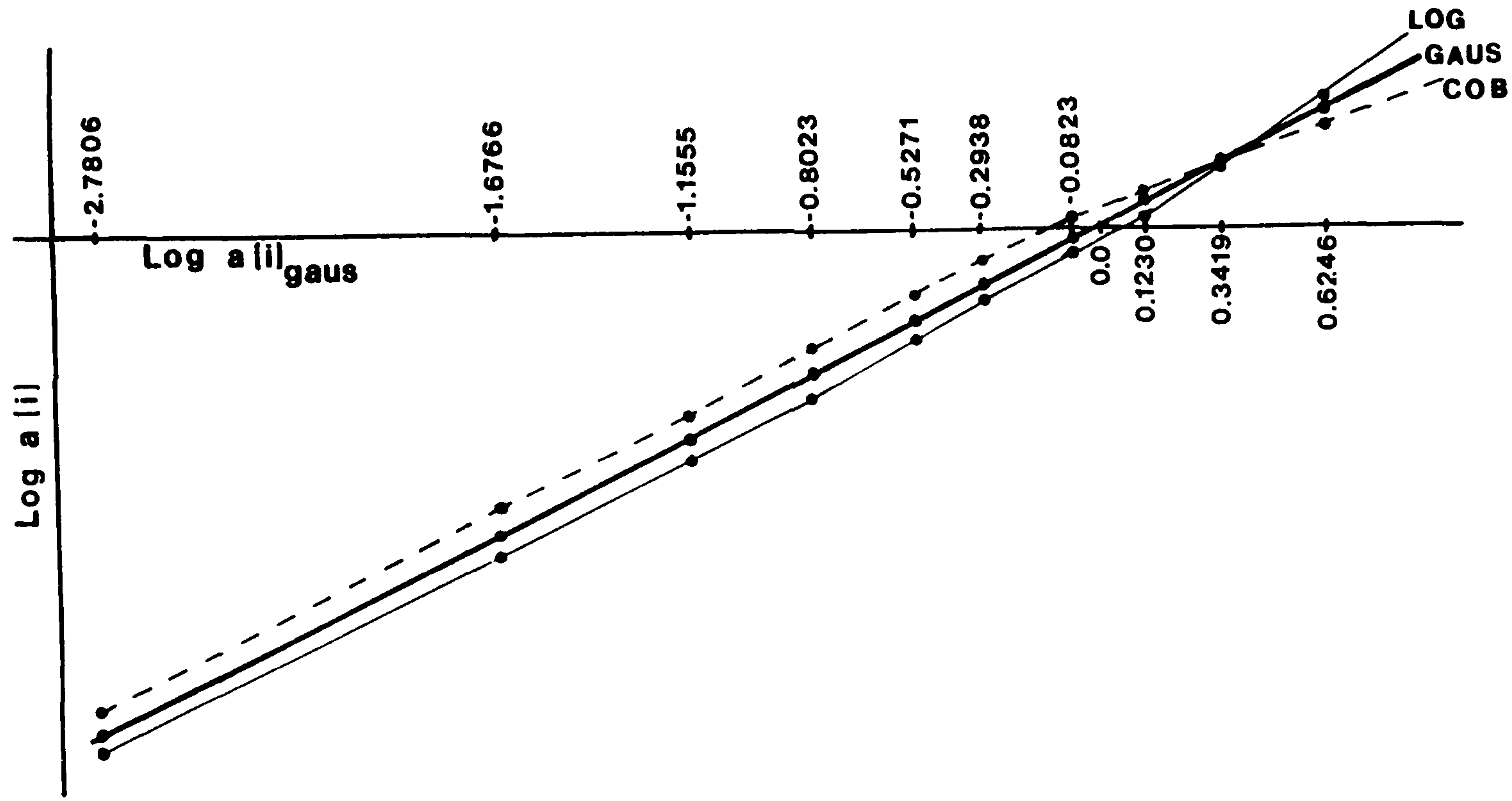

Figure 1: Log $a(i)_Q$ vs. log $a(i)_{Gaus}$ for Q=Gaussian, cosine-bell and logistic

In order to evaluate the performance of the pushback in its various forms, data were simulated from six situations. The situations and their densities for simple and compound situations are:

Gaussian (Gaus) $\frac{1}{\sqrt{2\pi}}\exp\{-(y)^2/2\} = G(0,1)$

slash (slash) = ratio of an independent Gaussian to unif(0,1)

$$\frac{1}{\sqrt{2\pi}y^2}\left[1-\exp\{-\tfrac{1}{2}y^2\}\right] \qquad \text{for } y \neq 0$$

$$\frac{1}{2\sqrt{2\pi}} \qquad \text{for } y = 0$$

Cauchy (Cauchy) $\frac{1}{\pi(1+y^2)}$

one wild Gaussian, scale 10 (OWG) $(n-1)G(0,1) + 1G(0,100)$

mixture, scale 10 (mix) $.95\, G(0,1)+.05G(0,100)$

slacu (slacu) = ratio of independent Gaussian to

$(\text{unif}(0,1))^{1/3}$

$$= \frac{6}{y^4\sqrt{2\pi}}(1-\exp\{-y^2/2\} - \frac{3}{\sqrt{2\pi}y^2}\exp\{-y^2/2\}$$

The six distributions above include the "three corners", the Gaussian, slash, and one-wild Gaussian. Each of these three represents one extreme of a data type we regard as likely to be encountered in practice. Gaussian data is considered to be extremely "nice" data and is the norm against which we judge data from other distributions, for example, saying data is from a distribution with "heavier tails" than the Gaussian. One-wild Gaussian (scale 10) data has one observation from a Gaussian distribution with mean zero and variance 100 and the rest from a standard Gaussian and so, is very likely to have one outlier. The slash density behaves as the Gaussian does in the middle and as the Cauchy does in the tails. The first part of this statement can be seen by looking at

$$\lim_{y \uparrow\downarrow 0} \frac{1}{\sqrt{2\pi}\, y^2}\left(1-\exp\left(-\tfrac{1}{2}y^2\right)\right)$$

which is, ignoring multiplicative constants, $\exp(-\frac{1}{2}y^2)$, the form of the Gaussian density. Similarly, the second part of the statement is seen by looking at

$$\lim_{y \downarrow -\infty} \text{ and } \lim_{y \uparrow -\infty} \left| \frac{1}{\sqrt{2\pi}\, y^2}\left(1-\exp\left(-\tfrac{1}{2}y^2\right)\right) \right| .$$

This shows that the slash tail-density decreases as y^{-2}, as does the Cauchy. Thus the slash satisfies the empirical observation that data is usually Gaussian in the middle, while having much longer tails than the Gaussian. The three corners are good quantitative standards against which to seek good performance.

The remaining three densities have been included to gain further information. The Cauchy is included in order to understand how the pushback works on data from a long-tailed but peaked distribution. We expect, however, to encounter Cauchy-like data rarely in practice, and, thus, do not seek good performance against the Cauchy. The mixture at scale 10, because it is a milder form of the one-wild Gaussian at scale 10, gives information on the transition from the Gaussian to the OWG. The tail-density of the slacu decreases as t^{-4} in comparison to the slash tail-density which decreases as t^{-2}. The slacu tail-density is like a t_3 tail-density. (Note that t_3 tends to behave like 15G(0,1) + 5G(0,9) (F. Hampel (1979))). These facts indicate that the slacu should give us information on the transition from G(0,1) to slash and, in a less clear way, on the transition from G(0,1) to a mixture-like density.

Cases for which the simulations were performed are listed in table 2. All are for sample size twenty. Variances were calculated using a swindle (G. Simon (1975)) and are based on 500 samples. The computations discussed here were done on the Statistics Department PDP11/40. Gaussian samples were generated via an algorithm due to Forsythe and Ahrens-Dieter (J. H. Ahrens and U. Dieter (1974)). Uniform random numbers were generated using Knuth's algorithm M (Knuth (1969)).

Table 2

Simulation Cases

```
  mjr-gaus-g;c(.8,1.0,1.2);;sl(.4,.8,1.0,1.2)
             m(.4,.8,1.0,1.2,2.0);w(.4,.8,1.0,1.2)
  mjr-log-g;c;sl;m;w(.8,1.0,1.2)
  mad-gaus-g;c;sl;m;w(.8,1.0,1.2),((.4,.8,1.2,2.0)/.6745)
  mad-log-          "                       "
  mjr-gaus*-g;c;sl;m;w(.4,.8,1.0,1.2,2.0)
m mjr-gaus**-       "           "
e mjr-log*-         "           "
d mjr-cob-g;c;sl;m;w;su(.4,.8,1.0)
i mjr-gaus***-      "            "
a 37.5%AD-gaus-g;c;sl;m;w;su(.4,.8,1.0,1.2)
n 45%AD-gaus-       "                  "
  55%AD-gaus-       "                  "
  70%AD-gaus-       "                  "
  75%AD-gaus-       "                  "
  80%AD-gaus-       "                  "
  85%AD-gaus-       "                  "
  90%AD-gaus-       "                  "

  mjr-gaus-g;c;sl;m;w;su(.4,.8,1.0,1.2,2.0)
  mjr-log-          "                  "
  mjr-gaus*-        "                  "
w mjr-gaus**-       "                  "
6 mjr-log*-         "                  "
' mad-gaus-g;c;sl;m;w((.4,.8,1.0,1.2,2.0)/.6745)
b mad-log-          "                  "
i mjr-cob-g;c;sl;m;w;su(.4,.8,1.0)
w mjr-gaus***-      "           "
e 37.5%AD-gaus-g;c;sl;m;w;su(.4,.8,1.0,1.2)
i 45%AD-gaus-       "                  "
g 55%AD-gaus-       "                  "
h 70%Ad-gaus-       "                  "
t 75%AD-gaus-       "                  "
  80%AD-gaus-       "                  "
  85%AD-gaus-       "                  "
  90%AD-gaus-       "                  "

9 mjr-gaus-g(.4,.8,1.0,1.2);c(.8,1.0,1.2);sl(.4,.8,1.0,1.2);
'           m;w(.8,1.0,1.2);su(.4,.8,1.0,1.2,2.0)
b mjr-log-g;c;sl;m;w(.8,1.0,1.2);su(.4,.8,1.0,1.2,2.0)
i mad-gaus-g;c;sl;m;w(.8,1.0,1.2)
w mad-log-          "           "
e mjr-gaus*-g;c;sl;m;w;su(.4,.8,1.0,1.2,2.0)
i mjr-gaus**-       "                  "
g mjr-log*-         "                  "
h
t
```

*1 smooth
**2 smooths
***pushback on inner 16 only

4. Discussion of the Simulation Results

The results of the simulation cases listed in table 2 are given in figures 2-9 and in tables 3 and 4. For example, the results of the simulations for the first line in table 2 are given in figure 2. The scale estimate is MJR, the {a(i)} are the central order-statistic values from a Gaussian, the location estimate is the median, and the simulations were done for various k-values for Gaussian, Cauchy, slash, mix, and OWG data distributions. For ease of notation, we will refer to this pushback form as the MJR-Gaus-pushback median. In general, we refer to a scale estimate - {a(i)} distribution-pushback location estimate, the k-value and data-distribution being specified when necessary.

In figure 2, the first number listed at each k-value, data-distribution combination is the variance of the MJR-Gaus-pushback-median. The variance of this variance estimate is shown in parentheses here as it is in the similar figures which follow. The data-distribution is represented by the angle of the arc on the set of concentric circles and the k value is indicated by the circle radius. The inner circle contains values for a non-pushback estimate. For example, .0731 is the estimated variance of the median when the data is Gaussian.

Figure 2 also gives the MJR-logistic-pushback median variances. These values are the second entry in the figure.

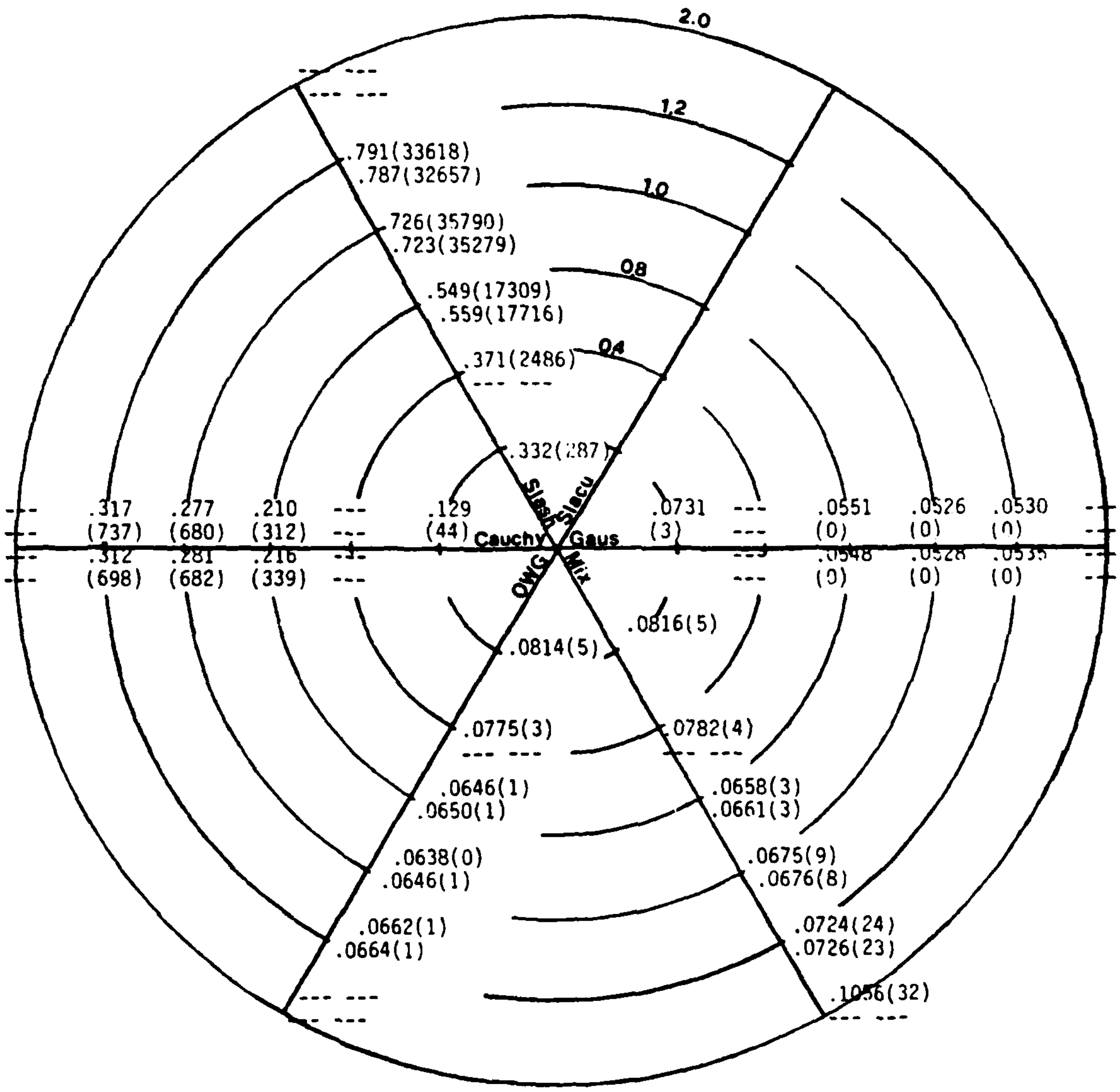

Figure 2: Variance and variance of variance x 10^6 (in parentheses) of MJR-Gaus-pushback median and MJR-log-pushback median

Similarly, the MJR-Gaus- and MJR-logistic-pushback w6-biweight and 9-biweight variances are given in figures 3 and 4, respectively.

Table 5 shows the optimum k-value (the k-value with minimum variance) and the corresponding optimum variance for a given scale estimate, {a(i)} distribution, location estimate and data distribution. From table 5 and figures 2-4, we see the following:

- First, for MJR and each of the median, w6-biweight, and 9-biweight the variance at the optimum k-value for the Gaussian is less than that at the optimum k-value for the logistic {a(i)}.

- Second, for each of the location estimates, the simulations indicate that the optimum k-values for a given data-distribution for the the logistic and Gaussian {a(i)} versions of the pushback with s = MJR are the same.

- Also, looking at lines 1, 15, and 29 of table 5, we see that the median and w6-biweight have less-changing optimum k values across distributions than the 9-biweight.

Approximately equality of k-values for the different distributions is of importance in its effect on the comparison of pushback and a known optimal estimate. We see this comparison in plots of the relative efficiency of pushback to a known good estimate, say w6-biweight, against the log

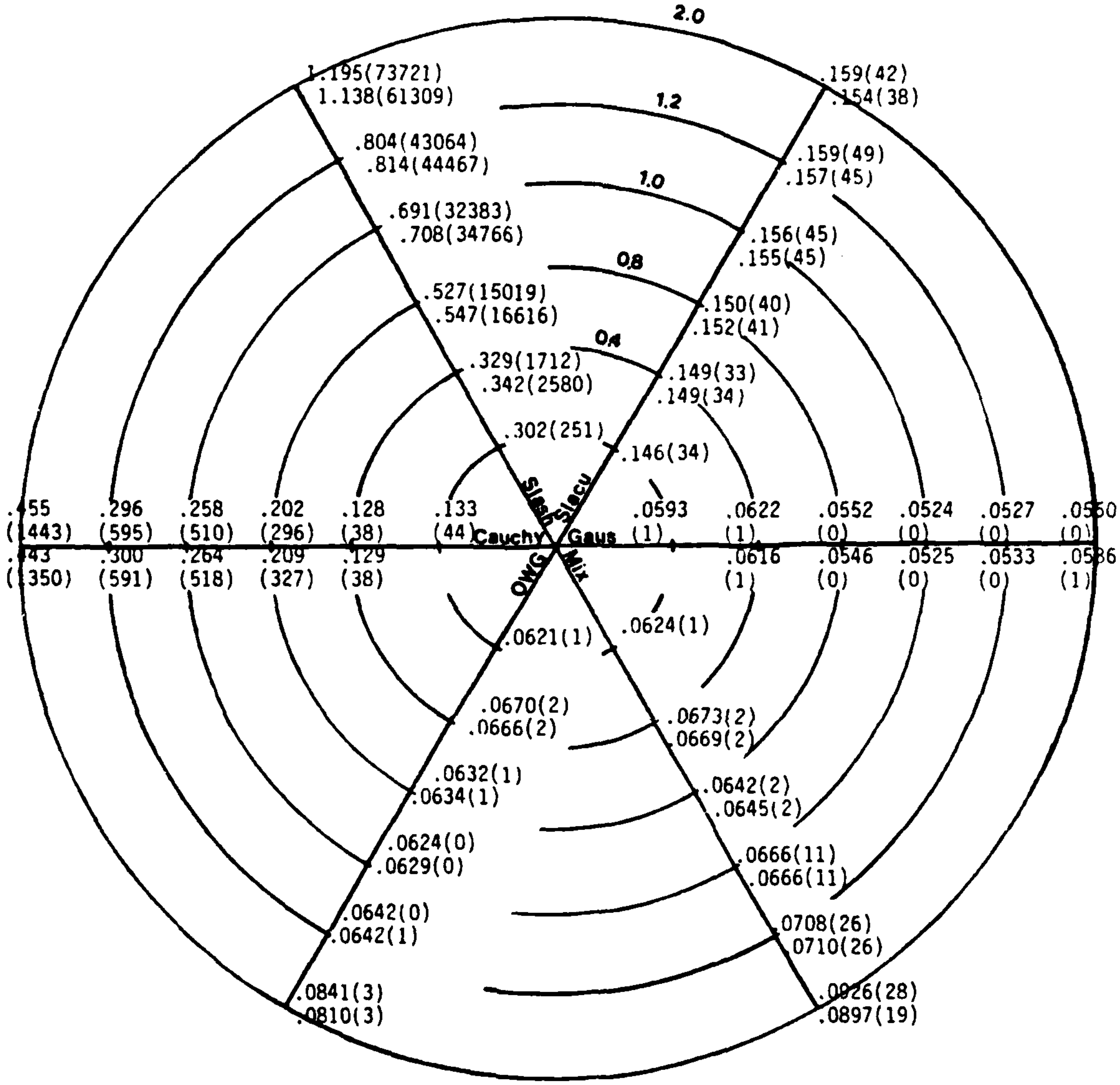

Figure 3: Variance and variance of variance x 10^6 (in parentheses) of MJR-Gaus-pushback w6-biweight and MJR-log-pushback w6-biweight

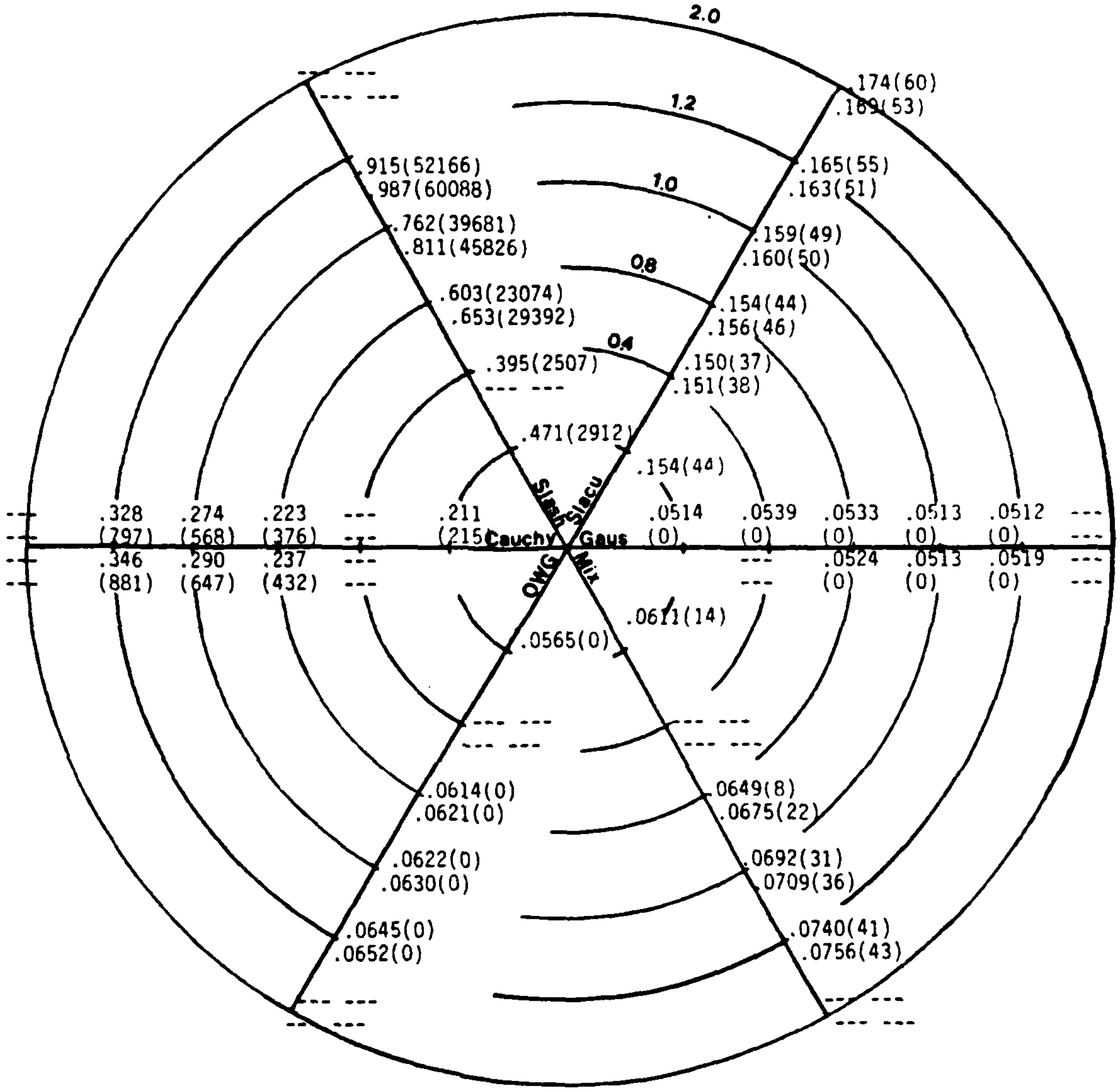

Figure 4 : Variance and variance of variance x 10^6 (in parentheses) of MJR-Gaus-pushback 9-biweight and MJR-log pushback 9-biweight

of the pushback constant, log k. Consider

$$E = \max_k\{\min_Q\{\text{rel eff}\}\} ,$$

where Q is the set of distributions specified, as a measure of the performance of the pushback. Approximately equal optimum k-values across distributions will help to keep E high by preventing the relative efficiency plot for one or more distributions from plunging while the other plots remain high.

Clearly, investigation into the effect of a particular scale estimate and {a(i)} distribution is necessary for tuning and understanding of the procedure. We first consider alternative scale estimates. Figure 5 shows the MAD-Gaus-pushback median variances and the MAD-logistic-pushback median variances, in that order. (Some of the k values shown here are approximate. Each of .4, .8, 1.0, 1.2, and 2.0 is divided by .6745 and the rounded values .6, 1.2, 1.5, 1.8, and 3.0 are shown in figure 5.) Figures 6 and 7 show similar results for w6-biweight and 9-biweight, with figure 6 again showing approximate k-values. From these values and those in table 5, we see again that the variances when Gaussian {a(i)} are used are better than or very close to the variances when logistic {a(i)} are used.

Given the results discussed above, that when s = MAD or MJR, and for the median and w6-biweight, Gaussian {a(i)} yield procedures which generally have smaller variances than

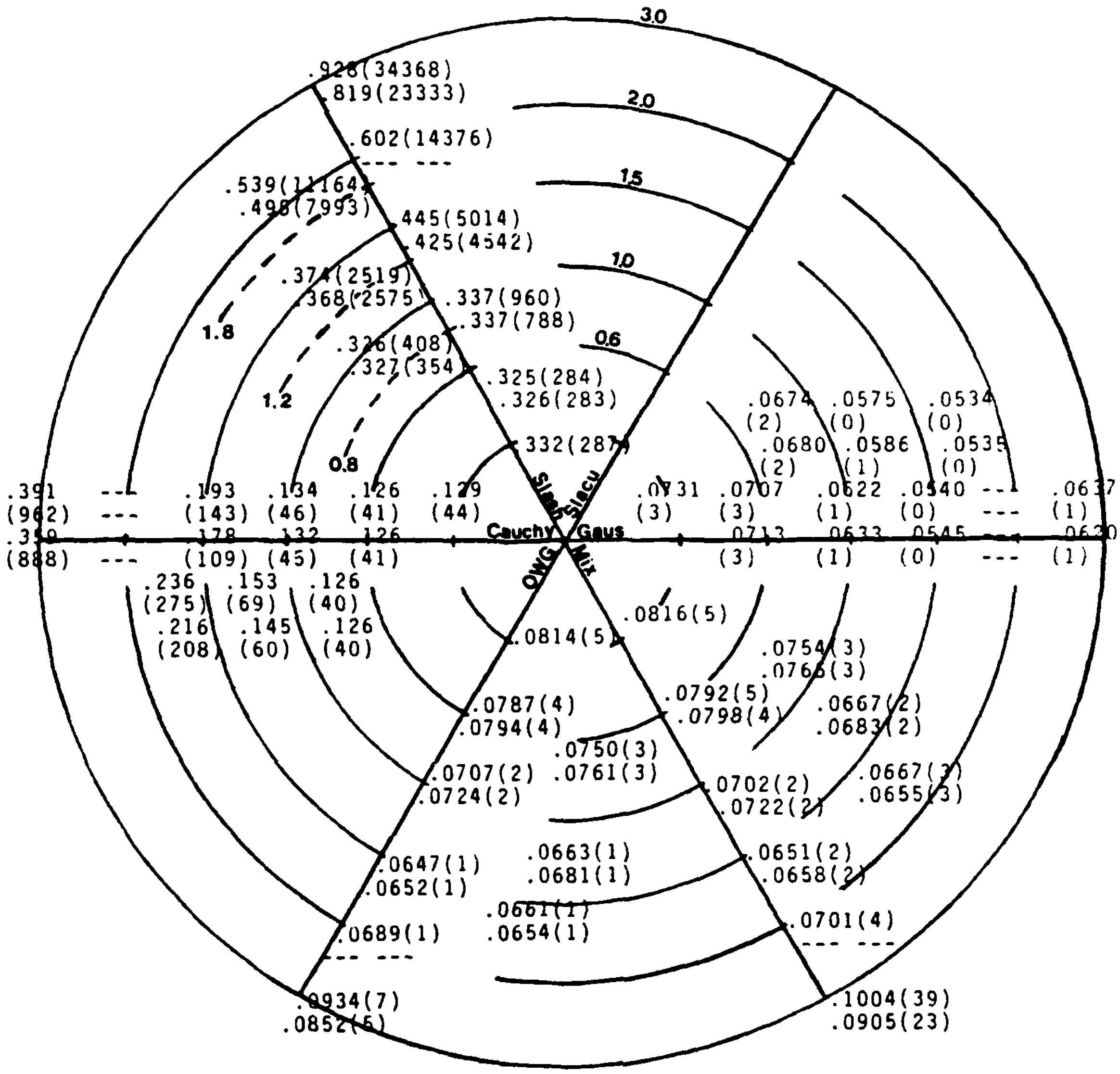

Figure 5: Variance and variance of variance x 10^{6} (in parentheses) of MAD-Gaus-pushback median and MAD-log-pushback median*

*k-values are approximate; see note in text.

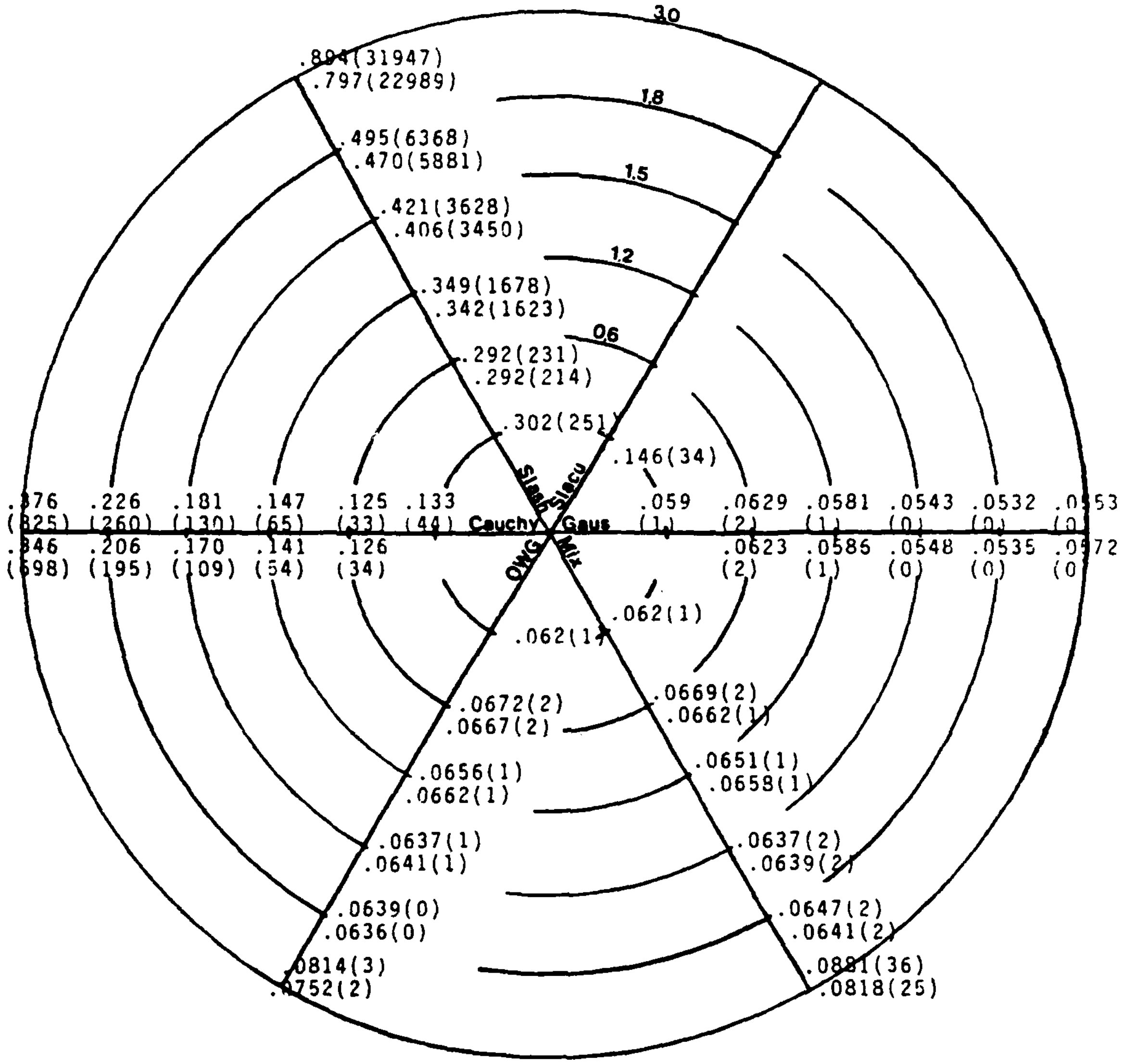

Figure 6: Variance and variance of variance x 10^5 (in parentheses) of MAD-Gaus-pushback w6-biweight and MAD-log-pushback w6-biweight*

*k-values are approximate; see note in text.

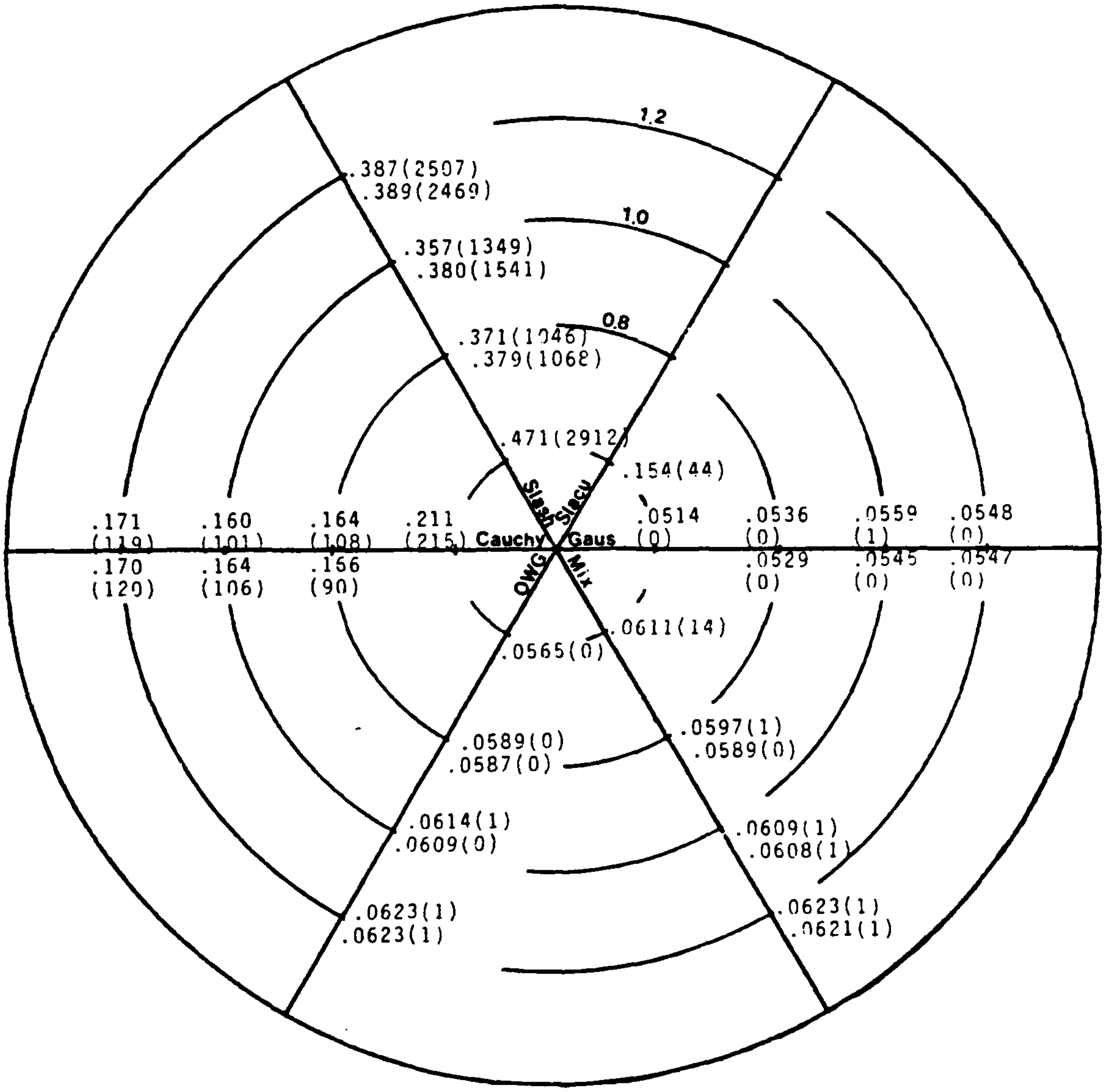

Figure 7: Variance and variance of variance x 10^6 (in parentheses) of MAD-Gaus-pushback 9-biweight and MAD-log-pushback 9-biweight

those in which logistic {a(i)} are used, we consider a third form of {a(i)}, cosine-bell {a(i)}. Figures 8 and 9 give results for the MJR-cob-pushback median and w6-biweight. Table 5 shows smaller variances for cosine-bell {a(i)} than for Gaussian {a(i)} at optimum k-values, but that the optimum k-values for the cosine-bell {a(i)} are more spread than those for Gaussian {a(i)}. Figures 10 and 11 show plots of the relative efficiency of the pushback to w6-biweight against the log pushback constant, ln k, for the MJR-Gaus-pushback median (Figure 10) and the MJR-cob-pushback median (Figure 11). Relative efficiency is again (variance w6-biweight)/ (variance pushback) for the particular form of the pushback being studied. The cosine-bell version has a slightly higher maximin relative efficiency (82% vs. 81%). In both cases, however, the slash curve keeps the values low by not attaining its maximum at a value of ln k near the maximizing ln k of the other distributions. Since cosine-bell and Gaussian {a(i)} have very close performance, we will continue testing with the more readily accepted norm, the Gaussian case.

Returning to our comparison of scale estimates, we look at the results for P%AD-Gaus-pushback median and P%AD-Gaus-pushback w6-biweight where P%AD is the scale estimate defined as the value at the lower P percent point of the absolute deviations from the median, {|y(i)-med{y(i)}|}. The P values used were 37.5, 45, 50 (i.e., the MAD), 55, 70, 75, 80, 85, and 90. The variances for these pushback forms

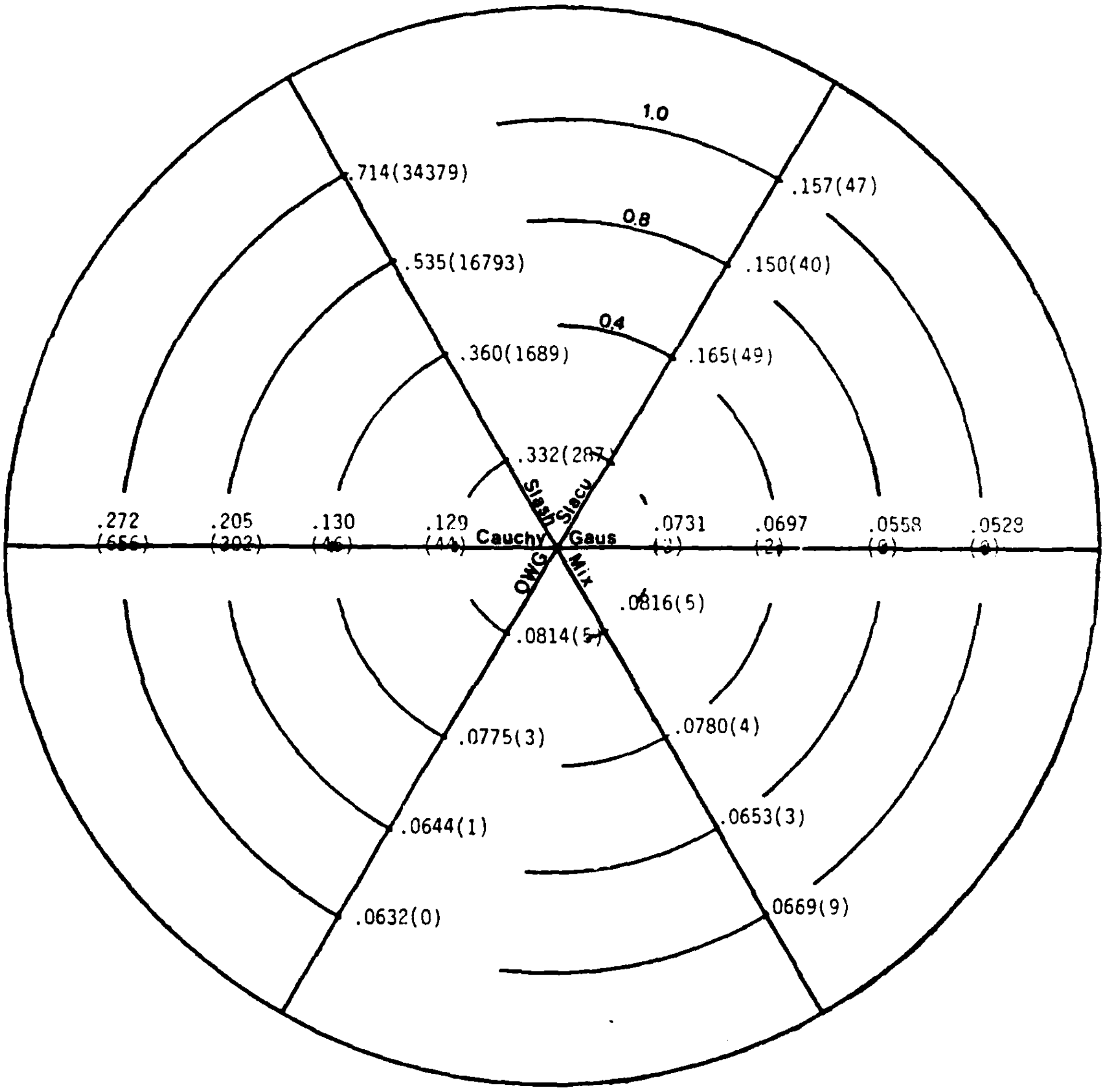

Figure 8: Variance and variance of variance x 10^6 (in parentheses) of MJR-cob-pushback median

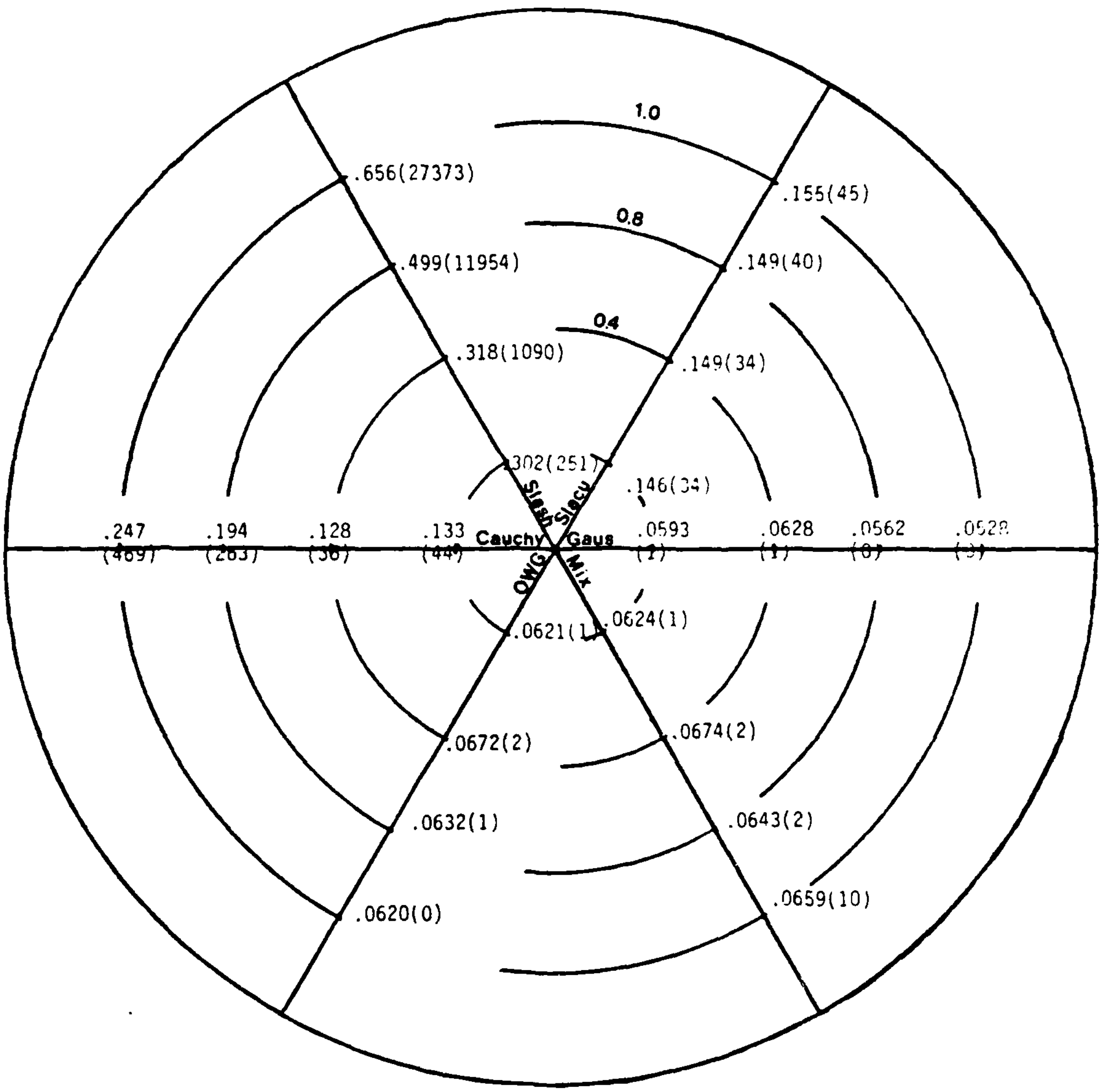

Figure 9: Variance and variance of variance x 10^6 (in parentheses) of MJR-cob-pushback w6-biweight

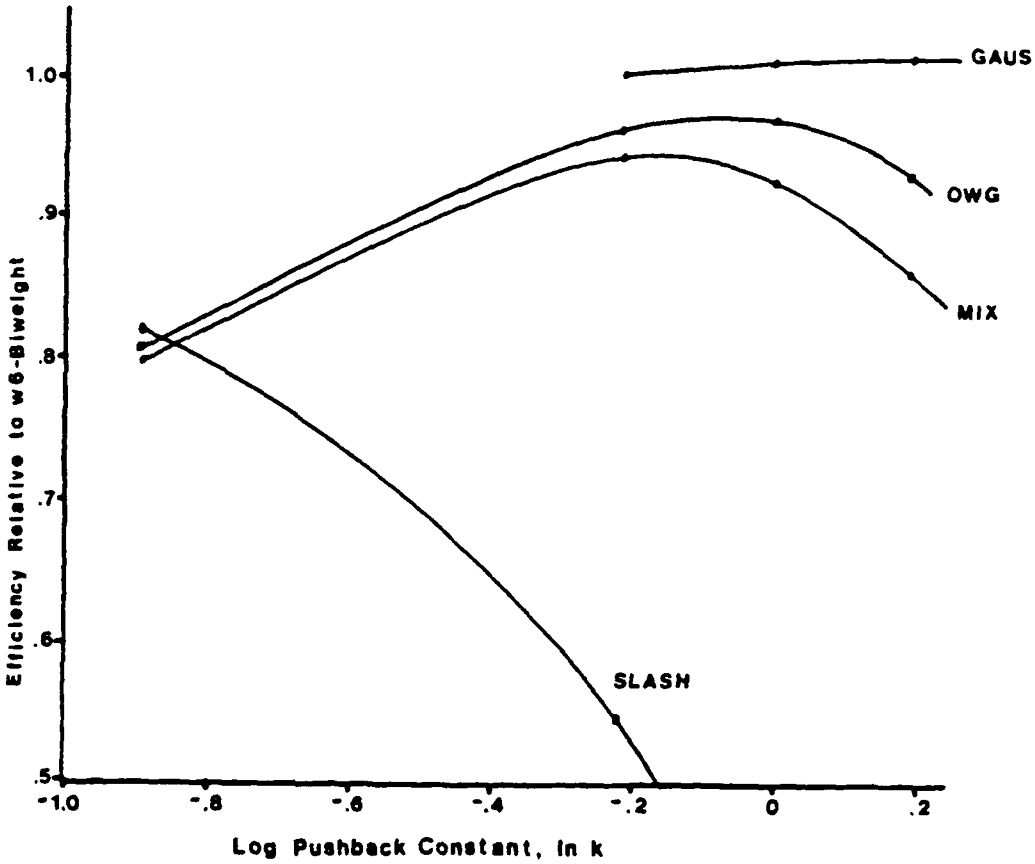

Figure 10 : Efficiency (relative to w6-biweight) of MJR-Gaus-pushback median vs. ln k

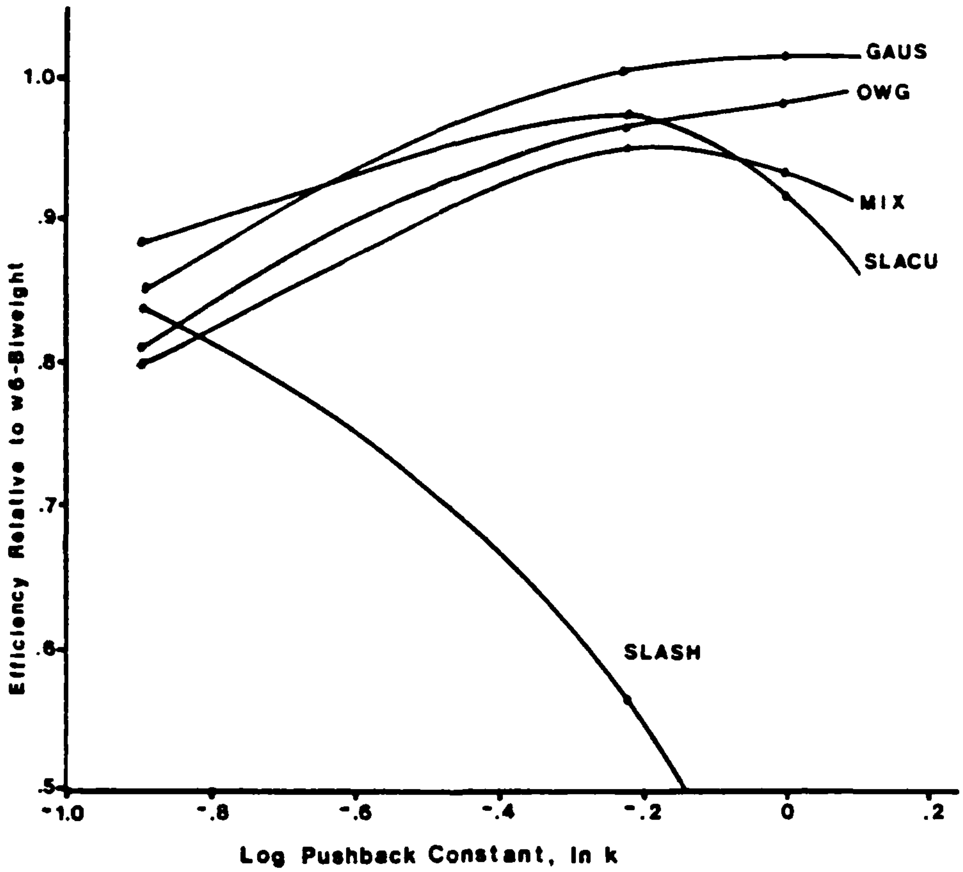

Figure 11: Efficiency (relative to w6-biweight) of MJR-cob-pushback median vs. ln k

are shown in tables 3 and 4 for the P%AD-Gaus-pushback median and P%AD-Gaus-pushback w6-biweight, respectively. Figures 12-20 show relative efficiency plots for the P%AD-Gaus-pushback median against ln k, where relative efficiency is again defined with respect to the w6-biweight. The following table shows the values of $\max_k\{\min_Q(\text{rel eff})\}$ for the various values of P used in P%AD.

P used in P%AD as scale estimate	max{min(rel eff)}
37.5	80
45	85
50	88
55	89
70	83
75	83
80	72
85	54
90	-

For each pushback listed in the table, Gaussian {a(i)} were used. We see that the pushback with s = 55%AD performs well in comparison to w6-biweight, achieving 89% or higher relative efficiency for all distributions considered.

Table 3

Variances and Variances of Variances x10^{6} (in parentheses)
of P9AD-gaus-pushback median

k	P =37.5	P 45	P 55	P 70	P 75	P 80	P 85	P 90
owj								
.4	.0310(5)	.0809(5)	.0803(5)	.0773(3)	.0760(3)	.0740(2)	.0712(2)	.0578(1)
.8	.0799(5)	.0786(4)	.0733(2)	.0651(1)	.0539(0)	.0648(1)	.0672(1)	.0737(2)
1.0	.0785(4)	.0758(3)	.0688(1)	.0637(0)	.0653(1)	.0691(1)	.0763(3)	.0859(6)
1.2	.0767(4)	.0723(2)	.0655(1)	.0672(1)	.0714(2)	.0789(4)	.0877(7)	.0982(1)
mix								
.4	.0914(5)	.0813(5)	.0806(5)	.0781(4)	.0769(3)	.0755(4)	.0725(3)	.0725(6)
.8	.0801(4)	.0785(4)	.0739(3)	.0554(2)	.0648(2)	.0707(22)	.0754(24)	.0922(60)
1.0	.0787(4)	.0751(3)	.0595(2)	.0657(2)	.0584(4)	.0753(17)	.0839(21)	.1005(44)
1.2	.0765(3)	.0717(3)	.0559(2)	.0693(4)	.0762(8)	.0850(20)	.0981(30)	.1100(39)
Gaussian								
.4	.0728(3)	.0726(3)	.0721(3)	.0700(2)	.0687(2)	.0670(2)	.0644(1)	.0615(1)
.8	.0717(3)	.0703(3)	.0560(2)	.0562(0)	.0536(0)	.0524(0)	.0523(0)	.0539(0)
1.0	.0702(3)	.0670(2)	.0605(1)	.0528(0)	.0523(0)	.0531(0)	.0564(0)	.0610(1)
1.2	.0682(2)	.0629(1)	.0561(0)	.0529(0)	.0543(0)	.0581(0)	.0633(1)	.0652(2)
slacu								
.4	.168(51)	.168(51)	.167(51)	.165(48)	.163(47)	.161(45)	.159(43)	.154(41)
.8	.167(51)	.164(49)	.161(47)	.153(43)	.155(45)	.159(51)	.163(50)	.151(44)
1.0	.165(49)	.160(46)	.154(46)	.155(46)	.159(50)	.160(46)	.164(54)	.165(52)
1.2	.163(48)	.156(44)	.151(40)	.159(49)	.163(54)	.165(54)	.164(53)	.168(53)
slash								
.4	.331(284)	.330(284)	.328(280)	.334(436)	.354(1321)	.419(6704)	.556(18593)	.941(41646)
.8	.327(269)	.329(369)	.325(417)	.465(5998)	.618(16641)	.886(37064)	1.187(63559)	1.519(57194)
1.0	.324(260)	.335(728)	.339(1024)	.609(15025)	.737(21476)	1.020(51789)	1.259(60155)	1.958(128090)
1.2	.323(265)	.354(1821)	.384(2697)	.680(17806)	.854(35041)	1.206(71279)	1.448(65919)	2.454(249576)
Cauchy								
.4	.128(43)	.128(43)	.127(41)	.130(45)	.138(51)	.169(122)	.226(391)	.382(1803)
.8	.127(41)	.125(40)	.129(41)	.200(226)	.273(50)	.389(1224)	.610(7416)	.873(23394)
1.0	.126(41)	.126(40)	.141(55)	.264(353)	.343(86)	.424(1431)	.539(8647)	.826(18751)
1.2	.126(41)	.131(44)	.168(94)	.300(462)	.374(94)	.469(1658)	.770(16300)	1.113(32046)

Table 4

Variances and Variances of Variances x10^6 (in parentheses)
of P%AD-gaus-pushback w6-biweight

k	P =37.5	P =45	P =55	P =70	P =75	P =90	P =95	P =99
owg								
.4	.0636(1)	.0642(1)	.0557(2)	.0655(2)	.0663(1)	.0655(1)	.0646(1)	.0631(1)
.8	.0657(2)	.0671(2)	.0675(2)	.0529(1)	.0619(0)	.0626(0)	.0643(0)	.0691(1)
1.0	.0657(2)	.0678(2)	.0652(1)	.0624(0)	.0631(1)	.0653(1)	.0693(1)	.0745(2)
1.2	.0675(2)	.0678(2)	.0548(1)	.0642(0)	.0657(2)	.0705(1)	.0755(2)	.0845(4)
mix								
.4	.0637(1)	.0642(1)	.0651(1)	.0667(1)	.0659(2)	.0673(2)	.0658(2)	.0675(5)
.8	.0558(1)	.0674(2)	.0683(2)	.0629(1)	.0625(1)	.0587(27)	.0721(28)	.0839(55)
1.0	.0558(2)	.0677(2)	.0652(1)	.0535(2)	.0652(2)	.0724(24)	.0782(27)	.0951(45)
1.2	.0673(2)	.0671(2)	.0541(1)	.0562(3)	.0710(6)	.0778(18)	.0868(25)	.1023(40)
Gaussian								
.4	.0506(1)	.0610(1)	.0617(2)	.0623(1)	.0620(1)	.0615(1)	.0602(1)	.0585(0)
.8	.0620(2)	.0630(2)	.0529(1)	.0557(0)	.0534(0)	.0520(0)	.0515(0)	.0525(0)
1.0	.0624(2)	.0630(2)	.0602(1)	.0528(0)	.0520(0)	.0523(0)	.0539(0)	.0555(0)
1.2	.0625(2)	.0614(1)	.0566(0)	.0523(0)	.0530(0)	.0546(0)	.0550(0)	.0563(0)
slacu								
.4	.146(33)	.147(33)	.147(33)	.148(33)	.148(33)	.148(34)	.149(35)	.148(36)
.8	.148(32)	.149(33)	.151(38)	.149(39)	.151(40)	.155(45)	.160(47)	.160(46)
1.0	.149(32)	.149(33)	.150(40)	.153(43)	.156(46)	.159(48)	.160(50)	.159(43)
1.2	.150(35)	.149(38)	.148(37)	.157(48)	.150(50)	.159(46)	.158(42)	.159(42)
slash								
.4	.297(242)	.298(260)	.295(248)	.298(294)	.322(951)	.381(4513)	.527(16779)	.953(52306)
.8	.295(246)	.294(263)	.295(285)	.434(4769)	.563(11243)	.803(35706)	1.118(57735)	1.696(100565)
1.0	.294(227)	.304(556)	.318(795)	.548(9624)	.695(19826)	.967(53270)	1.274(76240)	2.043(161876)
1.2	.292(232)	.326(1263)	.352(1917)	.648(15593)	.828(34803)	1.135(66575)	1.476(90117)	2.442(255504)
Cauchy								
.4	.130(38)	.129(37)	.126(34)	.129(38)	.135(45)	.162(114)	.216(302)	.377(1707)
.8	.127(35)	.125(33)	.126(36)	.189(180)	.251(39)	.339(819)	.525(4466)	.805(13170)
1.0	.126(35)	.125(33)	.137(47)	.242(295)	.311(59)	.399(1068)	.627(9052)	.967(22099)
1.2	.126(35)	.129(37)	.160(86)	.284(421)	.353(75)	.455(1527)	.789(21571)	1.212(38547)

Table 5

Optimum k Value and Corresponding Variance: variance (optimum k value)

		pushback	Gaussian	mix	owg	Cauchy	slash	slacu
	1	mjr-gaus	.0526(1.0)	.0558(.0)	.0639(1.0)	≤.130(≤.4)	≤.371(≤.4)	---
	2	mjr-log	.0528(1.0)	.0651(.8)	.0646(1.0)	≤.130(≤.4)	≤.389(≤.4)	---
	3	mjr-cob	≤.0528(≥1.0)	.0653(.8)	≤.0632(≥1.0)	≤.130(≤.4)	≤.390(≤.4)	.150(.8)
	4	mad-gaus	.0534(1.8)	.0651(1.5)	.0647(1.5)	≤.126(≤.6)	≤.325(≤.6)	---
m	5	mad-log	.0535(1.8)	.0655(1.8)	.0652(1.5)	.126(.8)	≤.326(≤.6)	---
e	6	37.5%AD-gaus	≤.0682(≥1.2)	≤.0765(≥1.2)	≤.0767(≥1.2)	.126(1.8)	≤.323(≥1.2)	≤.163(≥1.2)
d	7	45%AD-gaus	≤.0629(≥1.2)	≤.0717(≥1.2)	≤.0723(≥1.2)	.125(.8)	.329(.8)	≤.156(≥1.2)
i	8	55%AD-gaus	≤.0551(≥1.2)	≤.0559(≥1.2)	≤.0655(≥1.2)	≤.127(≤.4)	.325(.8)	≤.151(≥1.2)
a	9	70%AD-gaus	.0528(1.0)	.0654(.8)	.0637(1.0)	≤.130(≤.4)	≤.334(≤.4)	.153(.8)
n	10	75%AD-gaus	.0523(1.0)	.0648(.8)	.0639(.8)	≤.139(≤.4)	≤.354(≤.4)	.155(.8)
	11	80%AD-gaus	.0524(.8)	.0707(.8)	.0648(.8)	≤.159(≤.4)	≤.419(≤.4)	.159(.8)
	12	85%AD-gaus	.0523(.8)	≤.0725(≤.4)	.0672(.8)	≤.226(≤.4)	≤.556(≤.4)	≤.158(≤.4)
	13	90%AD-gaus	.0539(.8)	≤.0725(≤.4)	≤.0678(≤.4)	≤.382(≤.4)	≤.941(≤.4)	≤.154(≤.4)
	14	non-pushback	.0731	.0815	.0814	.125	.332	---
	15	mjr-gaus	.0524(1.0)	.0642(.8)	.0624(1.0)	≤.128(≤.4)	≤.329(≤.4)	≤.149(≤.4)
w	16	mjr-log	.0525(1.0)	.0645(.8)	.0629(1.0)	≤.129(≤.4)	≤.342(≤.4)	≤.149(≤.4)
6	17	mjr-cob	≤.0528(≥1.0)	.0643(.8)	≤.0620(≥1.0)	≤.128(≤.4)	≤.318(≤.4)	≤.149(≤.4)
'	18	mad-gaus	.0532(1.8)	.0637(1.5)	.0637(1.5)	≤.125(≤.6)	≤.292(≤.6)	---
b	19	mad-log	.0535(1.8)	.0639(1.5)	.0635(1.8)	≤.126(≤.6)	≤.292(≤.6)	---
i	20	37.5%AD-gaus	≤.0606(≤.4)	≤.0637(≤.4)	≤.0636(≤.4)	≤.125(≥1.2)	.294(1.0)	≤.146(≤.4)
w	21	45%AD-gaus	≤.0610(≤.4)	≤.0642(≤.4)	≤.0642(≤.4)	.125(1.0)	.294(.8)	≤.147(≤.4)
e	22	55%AD-gaus	≤.0566(≥1.2)	≤.0641(≥1.2)	≤.0648(≥1.2)	≤.125(≤.4)	≤.295(≤.4)	≤.147(≤.4)
i	23	70%AD-gaus	≤.0523(≥1.2)	.0629(.8)	.0624(1.0)	≤.129(≤.4)	≤.298(≤.4)	≤.148(≤.4)
g	24	75%AD-gaus	.0520(1.0)	.0626(.8)	.0619(1.8)	≤.135(≤.4)	≤.322(≤.4)	≤.148(≤.4)
h	25	80%AD-gaus	.0520(.8)	≤.0673(≤.4)	.0626(.8)	≤.162(≤.4)	≤.381(≤.4)	≤.148(≤.4)
t	26	85%AD-gaus	.0515(.8)	≤.0658(≤.4)	.0643(.8)	≤.216(≤.4)	≤.527(≤.4)	≤.149(≤.4)
	27	90%AD-gaus	.0525(.8)	≤.0675(≤.4)	≤.0631(≤.4)	≤.377(≤.4)	≤.953(≤.4)	≤.148(≤.4)
	28	non-pushback	.0593	.0624	.0621	.133	.302	.146
9	29	mjr-gaus	≤.0509(≥2.0)	≤.0619(≤.4)	.0614(.8)	≤.153(≤.4)	≤.395(≤.4)	≤.150(≤.4)
'	30	mjr-log	≤.0514(≥2.0)	≤.0627(≤.4)	.0621(.8)	≤.165(≤.4)	≤.415(≤.4)	≤.151(≤.4)
b	31	mad-gaus		≤.0597(≤.8)	≤.0589(≤.8)	.160(1.0)	.357(1.0)	---
i	32	mad-log	≤.0529(≤.8)	≤.0589(≤.8)	≤.0587(≤.8)	.164(1.0)	.379(.8)	---
w	33	non-pushback	.0514	.0611	.0565	.211	.471	.154
t								
	34	mjr-gaus*	≤.0547(≥1.0)	≤.0638(≥1.0)	≤.0626(≥1.0)	≤.129(≤.4)	≤.356(≤.4)	.148(.8)
	35	mjr-gaus**	≤.0543(≥1.0)	≤.0638(≥1.0)	≤.0621(≥1.0)	≤.123(≤.4)	≤.331(≤.4)	.148(.8)

*pushback median - inner 16 only
**pushback w6-biweight - inner 16 only

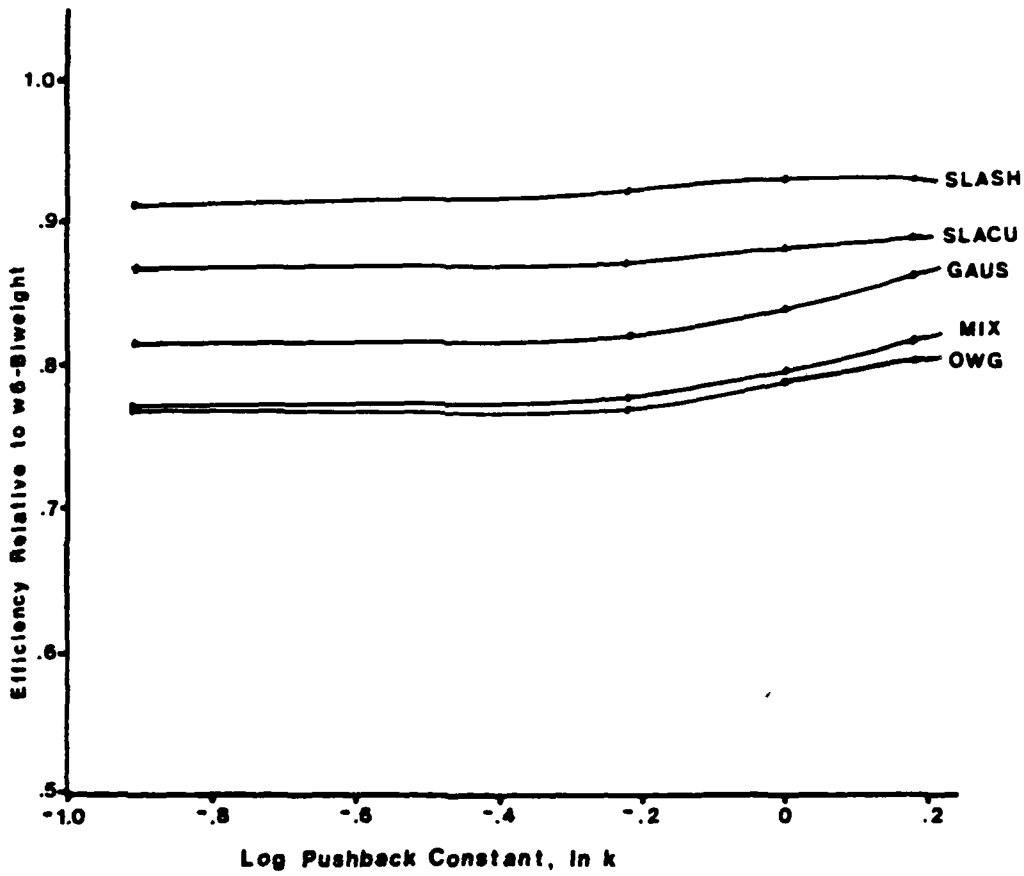

Figure 12 : Efficiency (relative to w6-biweight) of 37.5%AD-Gaus-pushback median vs. ln k

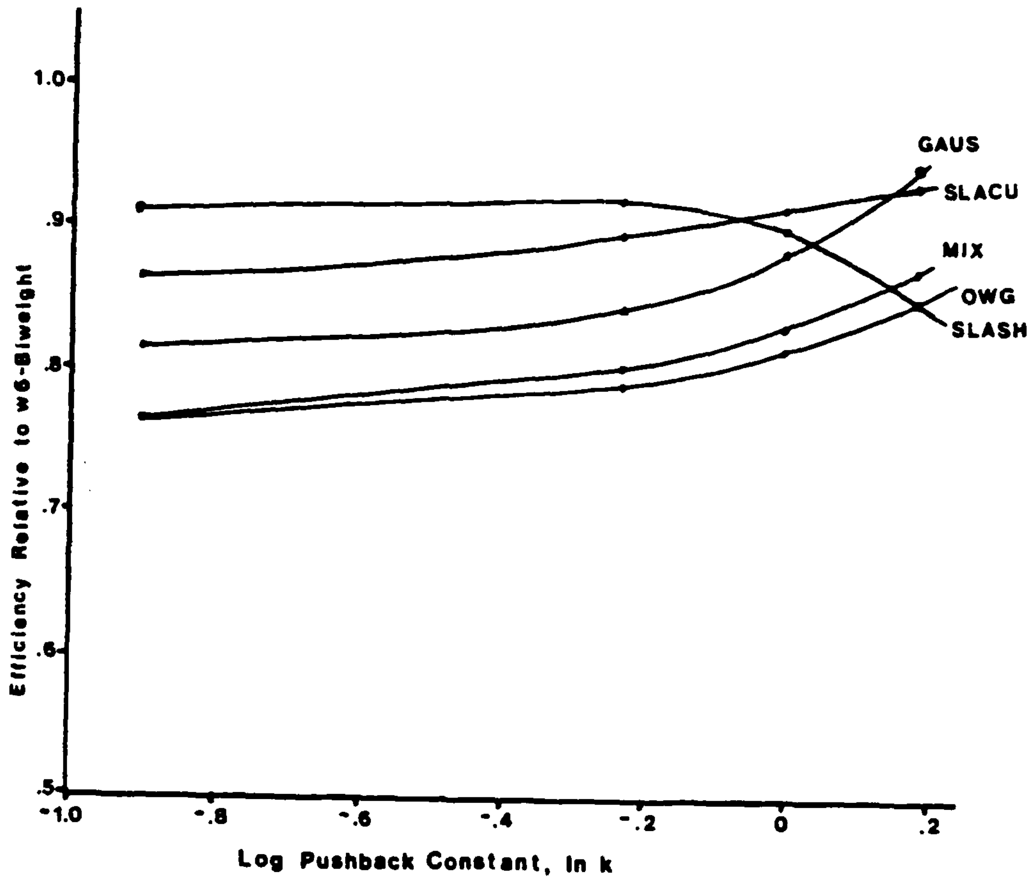

Figure 13: Efficiency (relative to w6-biweight) of 45%AD-Gaus-pushback median vs. ln k

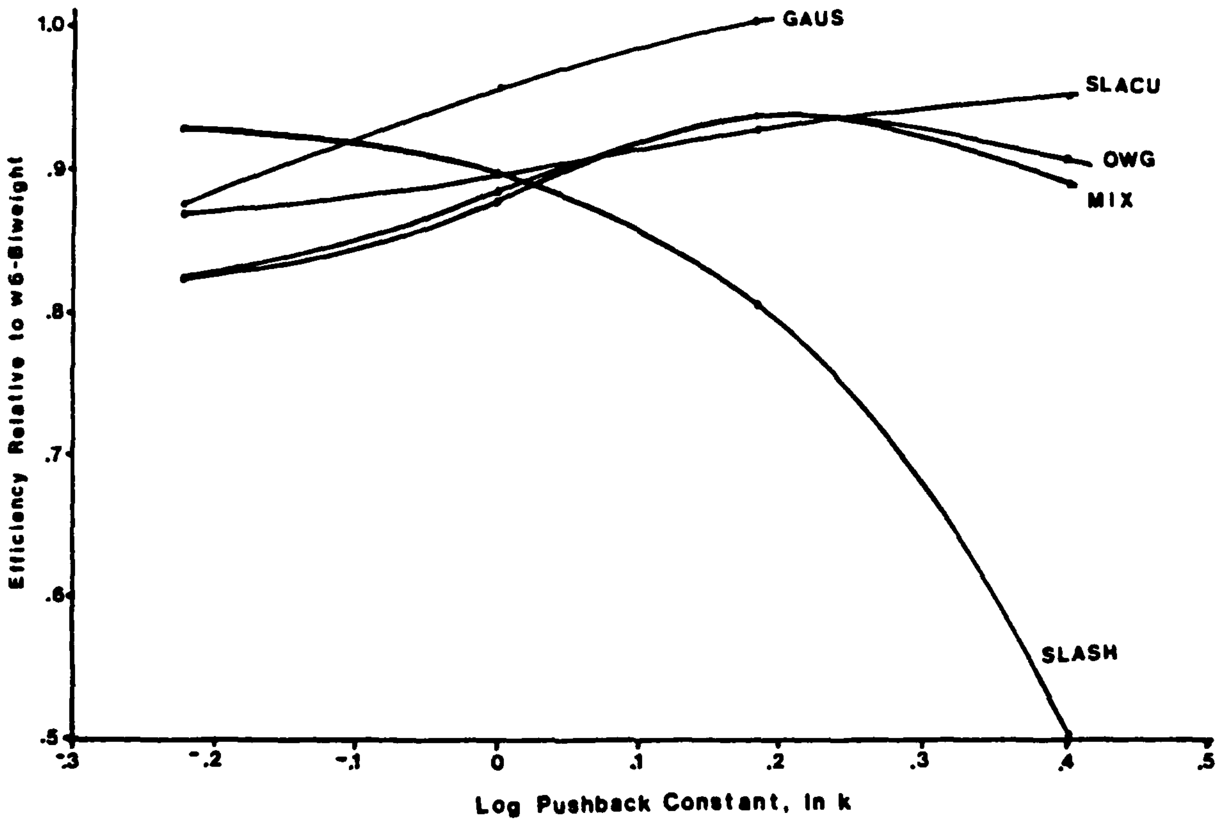

Figure 14 : Efficiency (relative to w6-biweight) of MAD-Gaus-pushback median vs. ln k

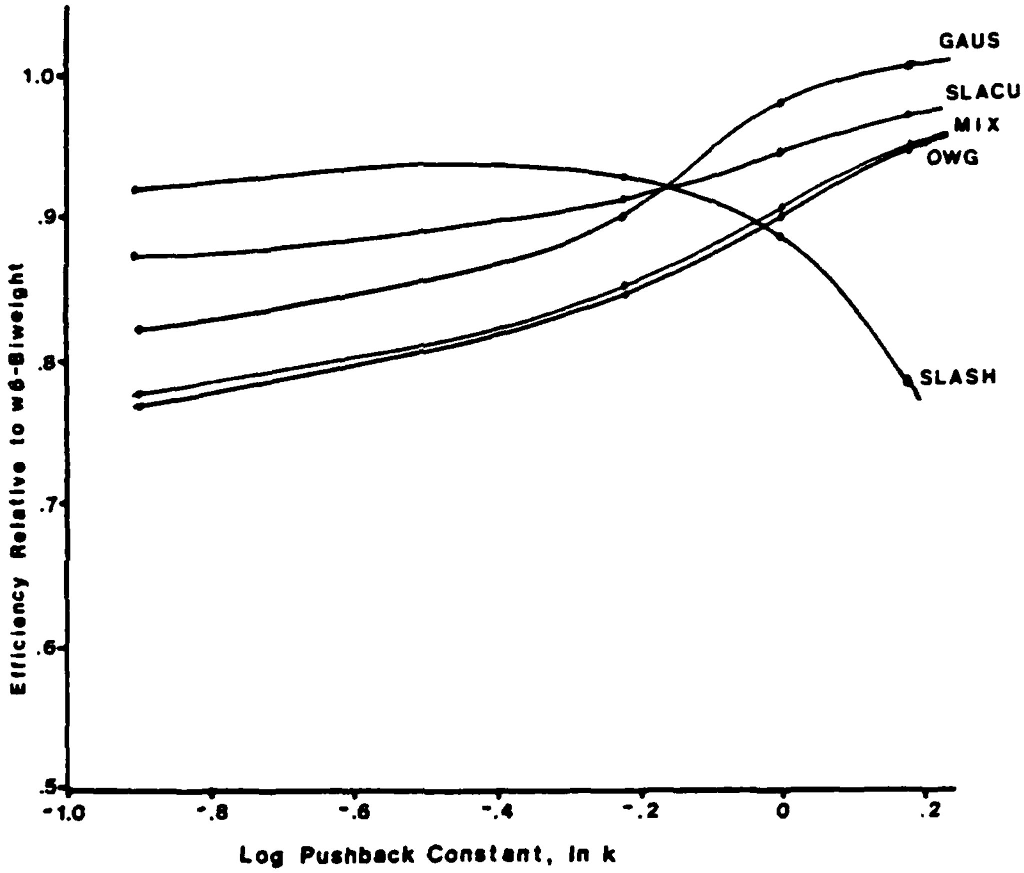

Figure 15 : Efficiency (relative to w6-biweight) of 55%AD-Gaus-pushback median vs. ln k

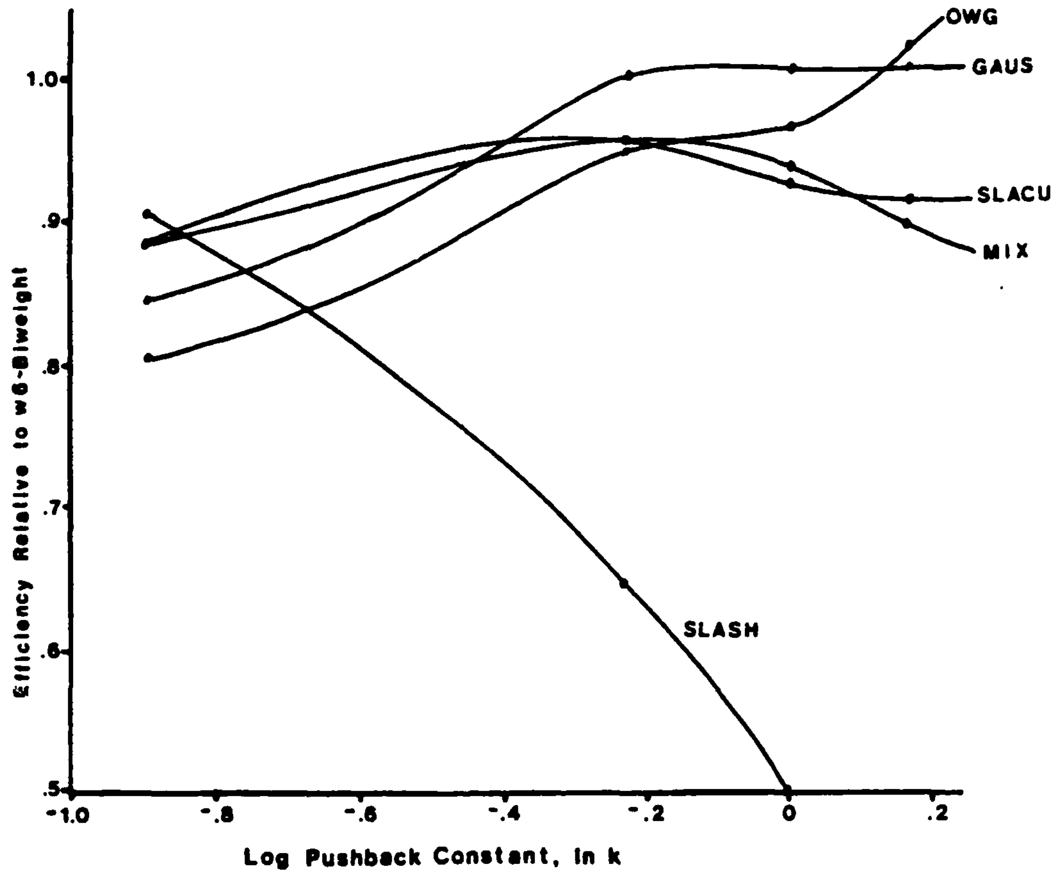

Figure 16 : Efficiency (relative to w6-biweight) of 70%AD-Gaus-pushback median vs. ln k

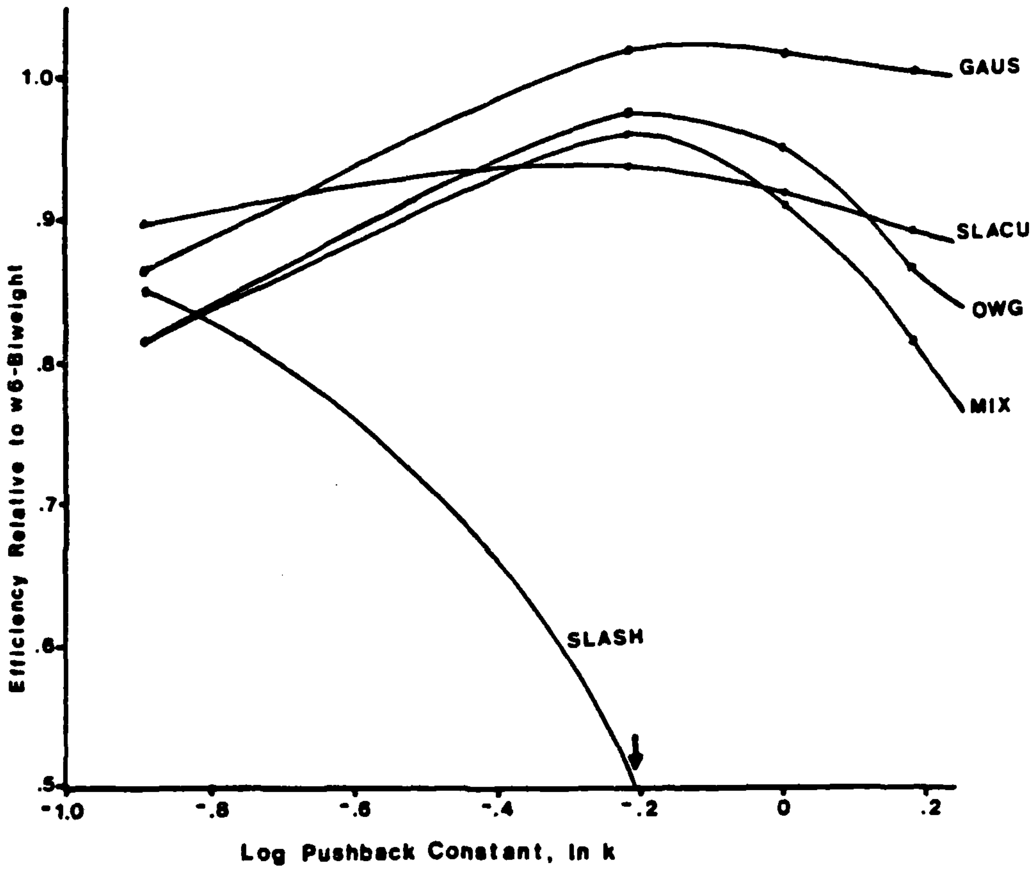

Figure 17: Efficiency (relative to w6-biweight) of 75%AD-Gaus-pushback median vs. ln k

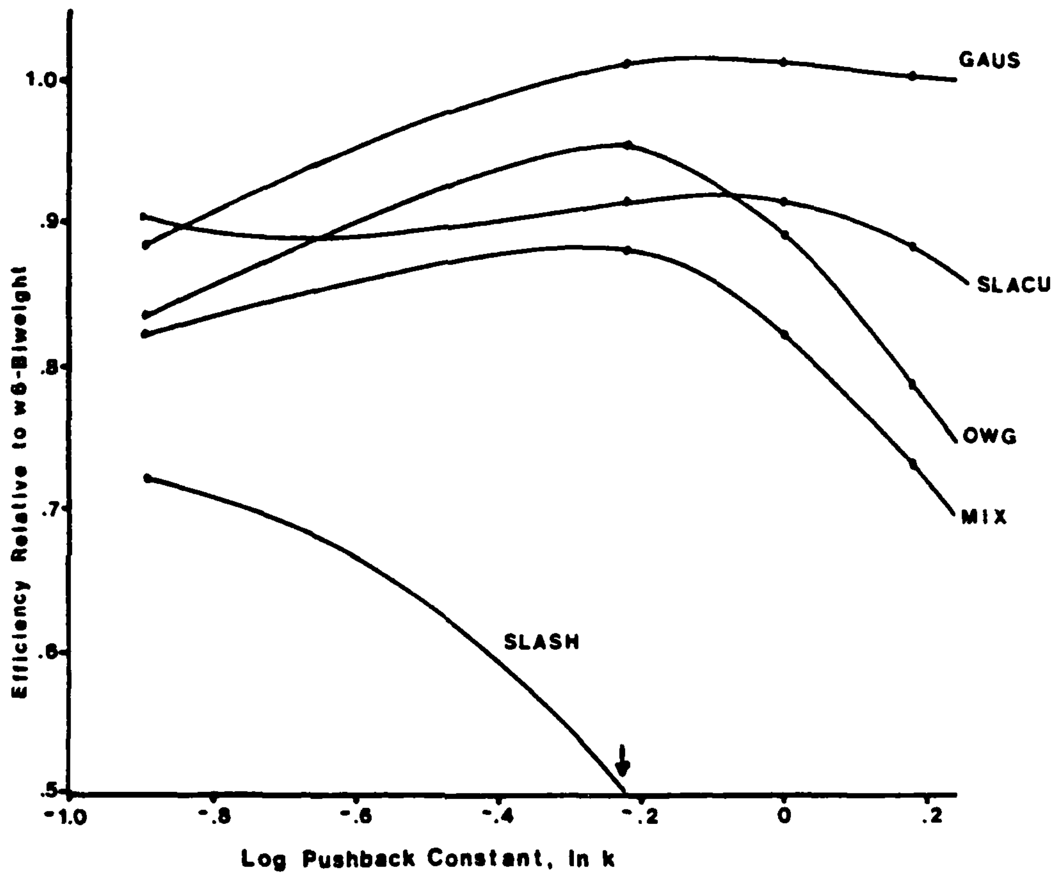

Figure 18 : Efficiency (relative to w6-biweight) of 80%AD-Gaus-pushback median vs. ln k

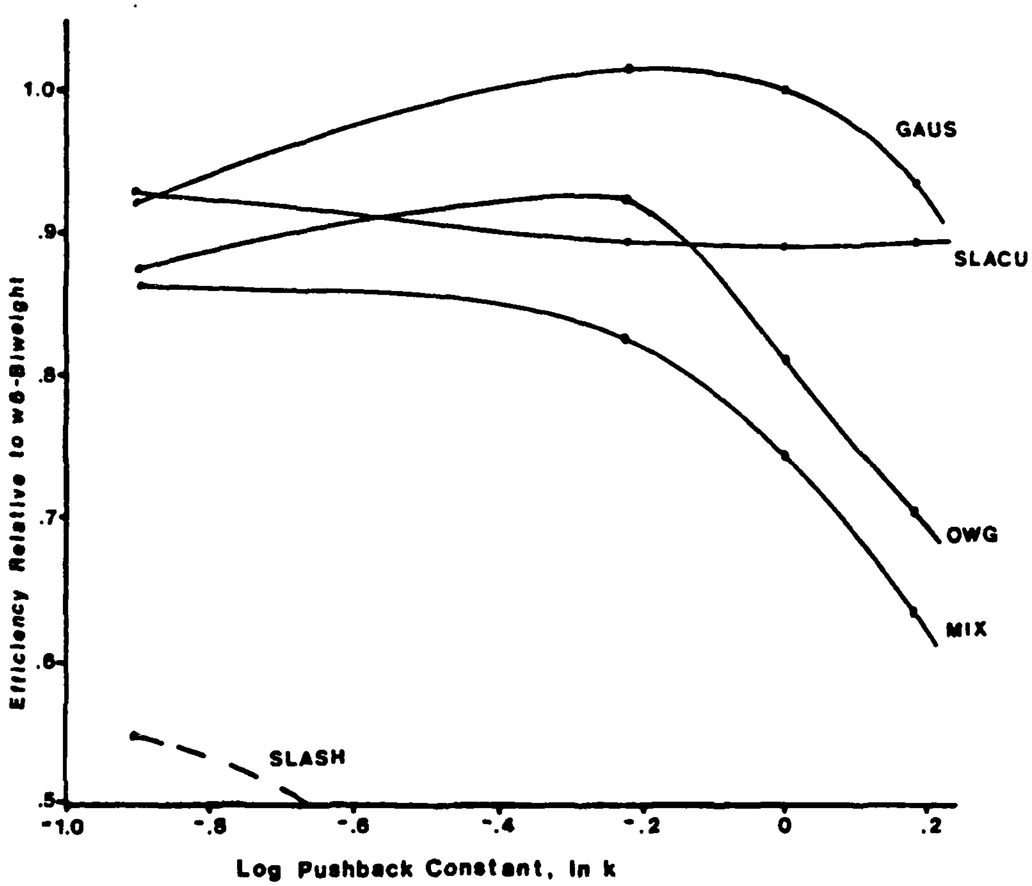

Figure 1[illegible]: Efficiency (relative to w6-biweight) of 85%AD-Gaus-pushback median vs. ln k

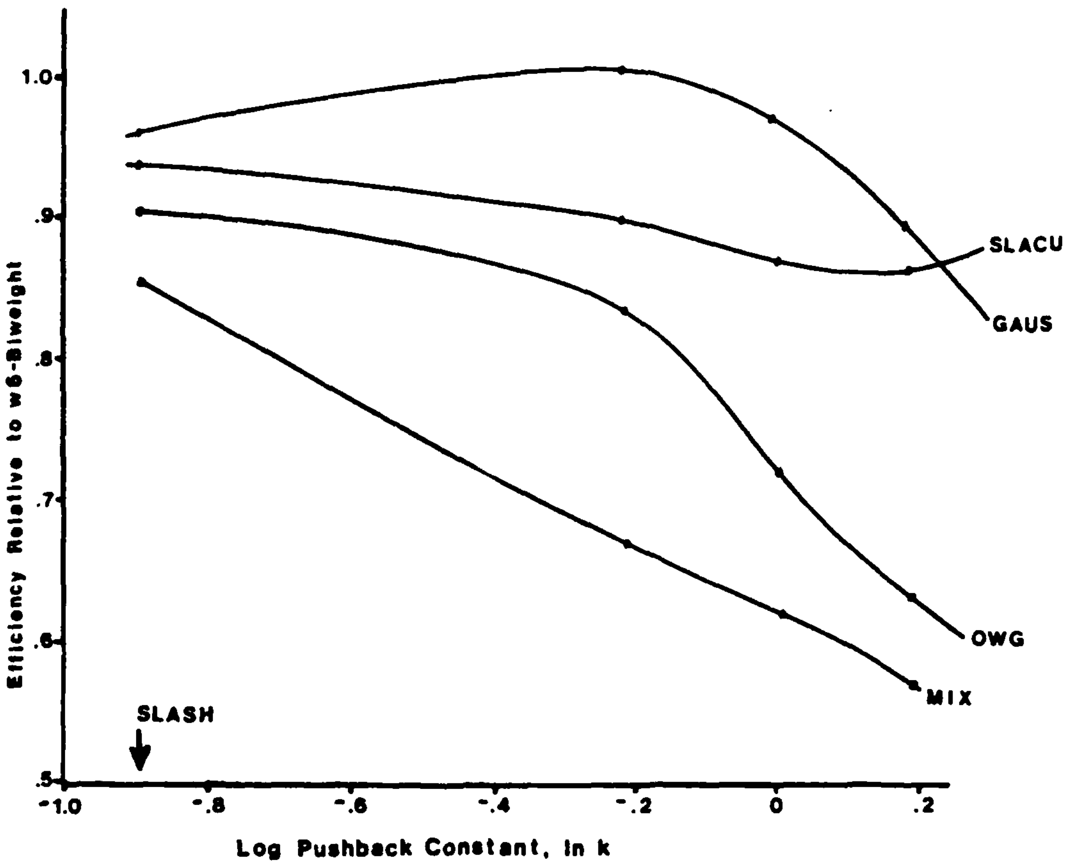

Figure 20 : Efficiency (relative to w6-biweight) of 90%AD-Gaus-pushback median vs. ln k

5. Conclusions from the Simulations

The results described in section 4 and the requirement of simplicity of the pushback procedure provide preliminary answers to the following question:

> What choices of
>
> - location estimate,
> - scale estimate, and
> - {a(i)} distrubution
>
> yield good performance for the associated pushback procedure while still maintaining its simplicity?

Gaussian {a(i)} were shown to perform better than logistic {a(i)} for s = MJR, and MAD with T = median, and w6-biweight. The Gaussian central order-statistic values also were shown to have performance very close to that of cob {a(i)} for s = MJR with T =median. We choose Gaussian {a(i)} as the central order-statistic values because of their performance and the wide acceptance of Gaussianity as the norm against which we judge other distributions.

The extensive simulations were done for T = median rather than T = w6-biweight or 9-biweight because of a preference for simplicity in procedure. Also, since the w6-biweight exhibits good performance, it is unlikely that modifications of it will have marked decreases in variance (see table 5).

The scale estimate s = 55%AD, when used with T = median and Gaussian {a(i)} achieves $\geq$ 89% of the efficiency of the w6-biweight (Figure 15). This relative efficiency is higher than those for the other scale choices tested when used with Gaussian {a(i)} and the median as the location estimate.

What the conclusions indicate then is, first, that Gaussian {a(i)} are a good form of central order-statistic values. This is because most data are Gaussian in the center. This choice makes it possible to find a line of the form $z(i) = k \cdot s \cdot a(i)$, where s = P%AD, which is parallel to the central P% of the data. The value of P will depend on the sampling situation. Second, we see that a good choice of scale estimate is the P%AD where *P*=55. *This choice is a* maxi-min choice since our criteria for performance is that the five distributions, Gaussian, OWG, mix, slash, and slacu, all exhibit good performance at the same value of k using this choice of s. The residuals of the data from the line specified by these choices of s and {a(i)} are then well-enough behaved for each of the five distributions that the media applied to these residuals, the pushback median, achieves good performance relative to the w6-biweight.

The simulation conclusions discussed thus far are based on overall descriptions of the behavior of the pushback procedures. Variances based on the 500 samples of size 20 for the various pushback procedures are compared as are the efficiencies relative to w6-biweight.

We would like to look at the behavior of a specific pushback for particular data configurations in order to both understand the procedure and to fine-tune the procedure. To do this we need the optimum estimate for the particular data configuration. Then identifying data configurations where the pushback performs poorly in comparison to the optimum estimate may indicate the tuning necessary to improve the performance of the procedure. We would also like to determine the minimum attainable variance for a particular sampling situation and the maximum attainable polyefficiency over several sampling situations in order to determine efficiencies relative to this optimum. Configural sampling and configural polysampling provide a method for achieving the analysis discussed in this paragraph. The post-configural polysampling pushback results are discussed in (Krystinik (1981)).

REFERENCES

Ahrens, J.H. and Dieter, U. (1974). "Computer methods for sampling from gamma beta, poisson and binomial distributions, Computing, 12, 225-227.

Birnbaum, A. and Dudman, J. (1963). "Logistic order statistics," Annals of Mathematical Statistics, 34, 658-663.

Bruce, A., Pregibon, D., and Tukey, J.W. (1981). "The second representing function for compound situations," Technical Report No. 186, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey.

Hampel, F.R. (1979). Letter to J. W. Tukey on suggestions for additional possible corners.

Knuth, D. E.,(1969). The Art of Computer Programming Vol. 2: Seminumerical Algorithms, Addison Wesley, Reading, Massachusetts.

Krystinik, K. (1981). "Post-configural polysampling pushback results," Technical Report No. 211, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey.

Nanni, Louis F. (1979). Unpublished work in the pushback.

Owen, D.B. (1962). _Handbook of Statistical Tables_, Addison-Wesley, Reading, Massachusetts.

Simon, G. (1975). "Swindles to improve computer simulation, with application to the problems of appraising estimates of location and dispersion in univariate samples," Technical Report No. 91, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey.

END

DATE
FILMED

5-82

DTIC